Financial Market Prediction During the COVID-19 Pandemic Using Machine Learning

Patrick Jaquart[a], Fabian Furtwängler[a], Marlene Kirchenbauer[a], Christof Weinhardt[a] [a]Institute of Information Systems and Marketing, Karlsruhe Institute of Technology, Germany

Abstract

The study examines the predictability of S&P 500 stock movements during the COVID-19 pandemic. It presents a comparative analysis of random forest and logistic regression models and evaluates the importance of different COVID-19-related features and control variables for the prediction task. The results show that all examined models significantly outperform a random classifier at a significance level of 1%. The random forest and logistic regression forecasts increase significantly in accuracy when a feature set including COVID-19-related data is utilized compared to a benchmark feature set without COVID-19-related variables. The random forest model trained on the full features set yields the statistically most accurate market forecasts. A feature importance analysis of the most accurate model reveals that the predictive power is not concentrated on a single COVID-19-related feature type but spreads over multiple different features. Due to its empirical nature, this study represents a snapshot between July 2020 and December 2021. Hence, future research may examine its application in future market environments. The paper introduces a machine learning-based market prediction framework during the COVID-19 pandemic. The presented results suggest that COVID-19-related features improve market forecasts during the COVID-19 pandemic and should be included in contemporary asset pricing models. This paper presents a machine learning-based market prediction framework and sheds light on market predictability in the changing market environment of the COVID-19 pandemic.

Keywords: COVID-19, Financial Market Prediction, Asset Pricing, Machine Learning, Random Forest

Electronic copy available at: https://ssrn.com/abstract=4516371

1. Introduction

The COVID-19 pandemic, induced by the SARS-COV-2 virus [14], has major ongoing implications for societies [24, 26] and economies [20, 28] around the globe. As with the whole economy, international financial markets have experienced various effects of the pandemic. Amongst these effects are more pronounced behavioral investment biases [23], increased investor fear and panic [18] and an overall increased market volatility [32, 27]. For the first months of the pandemic, researchers have documented a robust negative relationship between COVID-19 case numbers and financial market performance. These first months of the pandemic can be characterized by rapidly rising global COVID-19 case numbers [21], increased panic amongst investors [18], and mostly negative stock market developments around the globe. While these early pandemic months have been examined by various researchers that focus on specific pandemic aspects (e.g., see Ashraf [1], Ciner [7], Corbet et al. [8]), there remain open questions regarding the utilization of COVID-19-related data for financial market prediction in a later, more stable pandemic stage. Against this backdrop, we aim to tackle some of these questions by examining the predictability of S&P 500 stock price movements with COVID-19-related data. As many potential pandemic-related variables exist, we apply flexible machine learning models to shed light on this prediction problem. To summarize, our overarching research question is:

RQ: *What is the performance of machine learning models predicting stock price movements of the S&P 500 during the COVID-19 pandemic?*

We answer this research question by applying different random forest models to the task of predicting daily stock price movements of S&P 500 constituent stocks. We select this machine learning method as random forests can incorporate a large number of predictive variables and have been applied successfully to other financial market prediction tasks in the past. We further make use of logistic regression models to provide a predictive performance benchmark. Our analysis utilizes a comprehensive feature set consisting of various COVID-19-related data and control variables. Subsequently, we analyze the feature importance of all utilized features. We further evaluate the predictive power of COVID-19-related variables by comparing the out-of-sample predictive performance of different models per method type, whereby one of these models relies on the full feature set to make predictions. In

Electronic copy available at: https://ssrn.com/abstract=4516371

contrast, the other model only utilizes data unrelated to the COVID-19 pandemic.

2. Related Work

Due to the recent nature of the COVID-19 pandemic, there is limited evidence regarding its impact on financial markets. As of the time of writing this paper, most researchers have focused on the first months of the pandemic. For instance, [32] examine the impact of google search volume related to health news on stock indices in the twenty countries with the most COVID-19 cases as of March 30, 2020. Using daily data from the beginning of January to the end of March 2020, they find that higher health news search volumes have a negative effect on stock returns. [37] analyze the impact of the COVID-19 pandemic on stocks in the solar energy sector. They find that the intensity of the COVID-19 pandemic, approximated by the number of confirmed cases and government response stringency, negatively affects solar energy stocks from December 31, 2019, to June 4, 2020. [1] sheds light on the moderating effect of national culture on the impact of the COVID-19 pandemic on stock markets. He utilizes daily data from January 22th until April 17 in 2020 and presents evidence for a higher negative reaction of stock markets in countries with pronounced national-level uncertainty aversion. [30] examine the impact of COVID-19-related information on stock returns at a US state level. Using data from January 22th to June 30, they find that next-day stock returns in a state are negatively related to the number of COVID-19 cases in that state. They show that this effect is smaller in states with more governmental support and with better medical resources. [33] investigate the effect of global fear on commodity returns up to five days in the future on daily data between March 11 to May 18, 2020. They show that commodity returns rise as fear related to COVID-19, approximated by the Global Fear Index, increases. [27] studies the effect of reported cases and deaths in different countries on the US stock market, using daily data from April 9, 2019, to April 9, 2020. Utilizing a GARCH model, he finds that only the COVID-19 cases in China are negatively related to US stock returns. Further, he applies a vector autoregression model, which also indicates that the reported deaths in France and Italy negatively impact stock returns and positively impact the volatility index VIX. [6] analyzes the relationship between news and stock market index returns in the US and major European countries. He applies a panel quantile regression framework to daily data

Electronic copy available at: https://ssrn.com/abstract=4516371

from February 3, 2020, to April 17, 2020, and finds that stock market indices and different pandemic-related news exhibit asymmetric relationships. [8] investigate the impact of the COVID-19 pandemic on companies whose brand is associated with the word Corona. They deploy a GARCH model to analyze a dataset ranging from March 11, 2019, to March 10, 2020, and present evidence for a negative knock-on effect of the pandemic on the stock performance of these companies. [35] construct a COVID-19 daily fear index based on Google search volume between March and August 2020. They find a robust negative relationship in-sample between fear and returns of major stock market indices of various countries up to five days in the future. [7] examines US stock market predictability in the early stage of the COVID-19 pandemic. Applying a LASSO approach to a daily data set from January 2 to April 16, 2020, he finds that high yield and investment-grade corporate bonds exhibit significant predictive power for the US stock market. [25] analyze the impact of different governmental responses of G7 countries to the COVID-19 pandemic. Investigating daily data between July 1, 2019, and April 16, 2020, they find that lockdowns travel bans, and stimulus packages positively affect stock returns in the considered countries.

3. Methodology

3.1. Data

In this study, we utilize daily data ranging from the beginning of July 2020 until the end of December 2021. Our asset universe consists of all stocks in the S&P 500, whereby we use the index composition as of September 1, 2021 (the start date of our test set) to prevent a survivorship bias in our data. For company-specific data, we rely on the Compustat database [36], from which we acquire daily closing price, daily total return factor, and the S&P industry sector code. Furthermore, we download the values for the factors *small-minus-big*, *high-minus-low*, *market*, *momentum* and the daily risk-free rate from Kenneth French's data library [16]. We obtain daily values for the number of PCR tests in the US and the number of COVID-19 cases, deaths, and recoveries in all countries from the COVID-19 data repository of the John Hopkins University [21]. From the *Our World in Data* repository [31] by the Oxford University, we acquire the daily values for the number of hospitalized COVID-19 patients placed in an intensive care unit, vaccinations against COVID-19, the COVID-19 reproduction rate, and government stringency, for each country. We obtain the daily values of

Electronic copy available at: https://ssrn.com/abstract=4516371

the infectious disease equity market volatility index based on the work of [2] from the website of the *Economic Policy Uncertainty* research team [11]. Additionally, we download the relative number of google search queries for the keyword 'coronavirus' from Google Trends and create a continuous trend time series by scaling the values relative to the training period. Finally, acquire information about announcements of governmental fiscal help related to the finance sector-related from the *worldbank* homepage [38]. We only include NYSE trading days into our data set, remove days with missing observations and use the UTC timezone to aggregate all data. For sources that do not provide the data in UTC, we shift the respective daily timestamp by one day into the future to ensure that the feature set at a certain point in time does not consist of future data.

3.2. Feature Engineering

To receive our final feature set, we transform the downloaded data as follows: We calculate the daily total excess returns for each company by scaling the relative change of the daily closing prices with the daily total return factor and subtracting the daily risk-free return. We obtain the beta values for a specific day and company by regressing the daily values of the four Carhart-factors [5], three of which are based on Fama and French [12]), on the daily excess return of that company for the last 250 trading days prior to the selected day. This procedure is repeated for each day in our data set to receive daily beta values for every company in the data set. We construct additional features that represent the recent feature history to enable the different models to utilize information about the feature development over time. Concretely, we use the following set of intervals to aggregate the feature history for a specific feature: (0, 1], (1, 5], (5, 10], (10, 20], (20, 30], (30, 40]. Every interval in the set denotes the days prior to the prediction point over which a feature is aggregated. For return-based features, we calculate multi-period excess returns over these intervals. We construct a binary dummy variable for fiscal support announcements for each interval that indicates whether additional fiscal support was announced over that time window. We calculate the absolute and relative growth over the respective intervals for all other features that vary over time. Finally, we create one-hot encoded dummy variables for the weekday and the S&P industry sector. Table 1 gives a summary over the utilized feature types.

Electronic copy available at: https://ssrn.com/abstract=4516371

Table 1: Overview of the utilized feature types

COVID-19-related Features	**Other Features**
Number of active COVID-19 cases	Weekday
Number of COVID-19 recoveries	S&P industry sector code
Number of COVID-19 related deaths	Beta small-minus-big
Number of officially conducted COVID-19 tests	Beta high-minus-low
People vaccinated against COVID-19	Beta market
COVID-19 reproduction rate	Beta momentum
Government response stringency	Daily excess return
Google search trends	
Infectious disease equity market volatility index	
Announcements of goverment fiscal help	
Number of intensive care unit COVID-19 patients	
COVID-19 reproduction rate	

3.3. Targets

We construct binary target variables based on the next-day excess returns of the different companies. Concretely, we use the median return of the training set to split the two different classes. Thereby, observations with a higher next-day excess return than the training set median are grouped into Class 0, and observations with a lower next-day excess return are grouped into Class 1. This procedure ensures a balanced training set, which is especially important for noisy prediction problems, such as financial market prediction tasks. An imbalanced training set might cause prediction models to learn to predict the majority class solely.

3.3.1. Training, Validation, and Test Set

The first 12 months of data, which incorporate 124,146 observations, constitute our set for model training. We use the subsequent two months of data, consisting of 20,461 observations, as our validation set for model tuning. The final four months of data, amounting to 40,811 observations, make up our test set used for an out-of-sample evaluation of model performance and feature importance.

3.4. Models

3.4.1. Random Forest

Random forests are an ensemble learning method consisting of multiple, not perfectly correlated decision trees [3, 4]. They can be used for classification and regression problems and have been applied successfully to various

Electronic copy available at: https://ssrn.com/abstract=4516371

financial market prediction tasks [22, 17]. Here, we use random forest classifiers consisting of 100 individual decision trees. Furthermore, we perform a grid search on the validation set to optimize the maximum individual decision tree depth and the number of features considered at each split in a tree. As our feature set utilizes a relatively high number of dummy variables, we also test a relatively high numbers of maximum features randomly considered per split. A higher number of features generally leads to more precise individual trees but a higher correlation amongst the different trees in the random forest. Based on the grid search on the validation set, we choose a parameterization with a tree depth of 5, whereby 40% of all features are considered per split for our final model that includes all features. For the model that is not trained on COVID-19-related features, the final model has an individual tree depth of two and considers the root of the total number of features at each split. The full parameter grid is depicted in Table 2. For the other parameters of our random forests, we use the default parameter values of the scikit-learn library [9].

Table 2: Parameter tuning grid

Model	**Parameter Tuning Grid**
Random Forest	Maximum depth: {1, 2, 3, 5, 10, 15, 20, None}
	Maximum Features per Split: {sqrt, log2, 0.2, 0.4, 0.6, 0.8, 1}
Logistic Regression	Penalty: {l2, l1, elasticnet, none}

3.4.2. Logistic Regression

We further utilize a logistic regression as a benchmark model, as logistic regression models have become a standard for classification tasks over the last decades [19]. Logistic regression models follow the same principles as linear regression models, with the difference of a binary target variable. For the applied logistic regression model, we tune the applied penalty term based on the validation set accuracy as detailed in Table 2. Multiple penalty specifications yield the same validation set accuracy for the logistic regression models trained on the complete feature set. In this case, we apply an l2-penalty, which represents the default value of the scikit-learn library [29]. For the logistic regression model only using the features not directly related to COVID-19 uses, the l1-penalty yields the highest validation set performance. We further train our logistic regression models with the 'saga' [9] solver, as it trains fast on larger datasets and supports all considered penalty values in our tuning grid.

Electronic copy available at: https://ssrn.com/abstract=4516371

3.5. Evaluation

We evaluate our predictive models using the accuracy measure. Additionally, we utilize Diebold-Mariano tests [10] in combination with the mean absolute error loss function to compare the accuracy of different model forecasts pairwise, similar to Fischer and Krauss [15]. Since we train our models on a balanced training set, we can further calculate the probability for a random classifier with an accuracy of 50% to achieve a given model's test set accuracy based on the binomial distribution:

$$X \sim B(n = \#\text{test}, p = 0.5, q = 0.5), \tag{1}$$

where #test ($\widehat{=}40,811$) denotes the number of test set observations. Additionally, we calculate the precision and recall measures and the F1-score for every model and target class to examine potential differences in the predictability of the two classes. Also, we evaluate the feature importance of our random forest model using the measure of permutation feature importance. We consecutively permute all features randomly and calculate the overall decrease in model performance. We scale the overall decreases in model performance so that the feature importance scores sum up to one. If the decrease in model performance is higher for a given permuted feature, the respective model relies more on that feature for correct predictions.

4. Results

4.1. Model Performance

Table 3 presents the predictive accuracy of the different prediction models. The random forest utilizing COVID-19-related data achieves an out-of-sample predictive accuracy of 53.84%, while the logistic regression benchmark model trained on the same data set achieves a predictive accuracy of 52.74%. In comparison, the random forest and logistic regression model, trained on the data set without COVID-19-related data, only exhibit a test set accuracy of 50.87% and 50.56%, respectively.

Table 3: Overview of the different model out-of-sample accuracy scores

	Random Forest		**Logistic Regression**	
	Full Feature Set	No Covid-Related Features	Full Feature Set	No Covid-Related Features
Accuracy	0.538409	0.508711	0.527358	0.505574

Electronic copy available at: https://ssrn.com/abstract=4516371

Table 4 presents the results of the Diebold-Mariano tests [10], which show that the forecasts by the random forest model trained on the full feature set are significantly more accurate than the forecasts of all other models at a significance level of 1%. Additionally, the logistic regression model trained on the full feature set yields significantly more accurate forecasts than the models only trained on data unrelated to COVID-19.

Table 4: Diebold-Mariano test p-values to reject the null hypothesis towards the alternative hypothesis that the forecast of model i on the test sample is more accurate than the forecast of model j

i \ j		**Random Forest**		**Logistic Regression**	
		Full Feature Set	No Covid-Related Features	Full Feature Set	No Covid-Related Features
Random Forest	Full Feature Set	-	0.000000	0.001792	0.000000
	No Covid-Related Features	1.000000	-	1.000000	0.328625
Logistic Regression	Full Feature Set	0.998208	0.000000	-	0.000000
	No Covid-Related Features	1.000000	0.671375	1.000000	-

As shown in Table 6, we find that all models make statistically viable predictions. Our random forest and logistic regression model trained on the full features set have respective probabilities of 1.7875E-55 and 1.1295e-29 for a true model accuracy of 50%. For the random forest and the logistic regression trained on the feature set without COVID-19 data, these probabilities are 8.0668e-06 and 3.1262e-04. Table 5 presents precision, recall and F1-scores of the different models. It shows that the predictions of the full-feature random forest model are mostly balanced over both classes, as precision and recall are over 50% for Class 0 and Class 1. In comparison, the best performing logistic regression model is more likely to correctly classify observations that belong in Class 0, as it exhibits a recall score of 77.92% for Class 0 and 25.76% for Class 1.

Table 5: Overview of the different models' precision, recall and F1-scores

		Random Forest		**Logistic Regression**	
		Full Feature Set	No Covid-Related Features	Full Feature Set	No Covid-Related Features
Class 0	**Precision**	0.555060	0.526326	0.529684	0.512789
	Recall	0.547812	0.513177	0.779276	0.907783
	F1-Score	0.551412	0.519668	0.630684	0.655371
Class 1	**Precision**	0.521000	0.490744	0.519914	0.426090
	Recall	0.528309	0.503913	0.256760	0.073541
	F1-Score	0.524629	0.497242	0.343755	0.125433

Electronic copy available at: https://ssrn.com/abstract=4516371

Table 6: Probabilities for the different prediction models to exhibit a true model accuracy of 50%

	Random Forest		**Logistic Regression**	
	Full Feature Set	No Covid-Related Features	Full Feature Set	No Covid-Related Features
Probability	1.787494E-55	8.066806e-06	1.129511e-29	3.126180e-04

4.2. Feature Importance

Figure 1 presents the permutation feature importance scores of the 30 most relevant features of the full-feature random forest model. It shows that the model generally relies on numerous features to make predictions, as individual feature importance scores do not exceed 2%. The most important features are the newspaper-based infectious disease equity market volatility tracker and features that indicate the state of the pandemic, such as the changes in COVID-19-related active cases and deaths, the number of conducted COVID-19 tests and the number of COVID-19 patients in intensive care units. Regarding the regional distinction of features, we find that features indicating the state of the pandemic in the US *(11 features)* are most common among the 30 most important features, followed by features that indicate the state of the pandemic in the EU *(five features)* and China *(four features)*, as well as an aggregated world-wide scope *(four features)*. Furthermore, the remaining features represent different temporal characteristics of the infectious disease equity market volatility index, which is based on US newspapers.

Electronic copy available at: https://ssrn.com/abstract=4516371

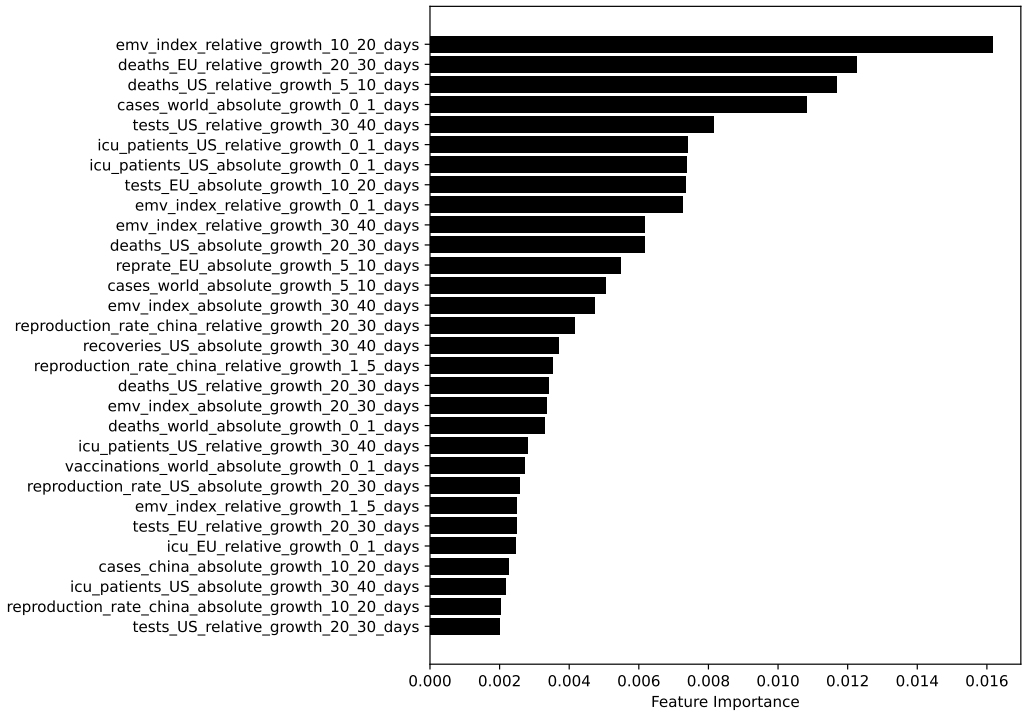

Figure 1: Permutation feature importance scores for the most relevant features of the full-feature random forest

5. Discussion

Our study demonstrates that COVID-19-related data exhibits predictive power for next-day price movements of S&P 500 constituent stocks. While previous studies have primarily examined the first months of the pandemic, a stage in the COVID-19 pandemic that can be characterized by investor fear [18] and high market volatility [32, 27], we shed light on a more extended, subsequent period. We document that the random forest model yields more accurate and balanced forecasts than the logistic regression model when training on the full feature set in our study. In combination with the finding that our grid search optimization results in medium-sized individual trees, this indicates that allowing for a limited number of high-dimensional feature interactions improves predictions for the problem at hand. The feature importance analysis reveals that the predictive power is not concentrated on a single COVID-19-related feature but is rather spread over multiple different feature types. The finding that US-based pandemic data tends to be

Electronic copy available at: https://ssrn.com/abstract=4516371

most relevant for making predictions might be explained by the fact that S&P 500 companies generate most of their revenue domestically [34]. An explicit limitation of this study stems from the dynamic evolution of the COVID-19 pandemic, as different mutations of the SARS-Cov-2 virus may become prevalent over time, and medical research will develop new methods for treatment and vaccination. While we aim to reduce part of this limitation by examining a relatively long period, our study still provides a snapshot in time, and similar analyses should be repeated for future stages of the pandemic. Furthermore, while we include well-established benchmark predictor variables [17] such as the three Fama-French factors [13], we can not entirely rule out a potential omitted variable bias. In this study, we evaluate market predictability through a binary classification task to create an overall benchmark in the academic literature in demonstrating the overall feasibility of financial market prediction using COVID-19 data. Based on this finding, future researchers could examine classification models with more than two classes or regression models trained to predict the exact excess returns of the different stocks. However, one has to keep in mind that this would further increase the complexity of the prediction problem and that one should generally expect relatively low predictability of developed financial markets in light of financial market efficiency. Future research could also examine the use of other machine learning methods and apply additional COVID-19-related features to the problem. While we have aimed to create a comprehensive feature set, new features might become available in the future due to new COVID-19-related developments.

6. Conclusion

Our study examines the predictability of S&P 500 stock movements during the COVID-19 pandemic utilizing COVID-19-related features. We find that a random forest and a logistic regression model trained on a comprehensive feature set that includes various COVID-19-related features yields more accurate predictions than their counterparts trained only on the subset of features unrelated to the COVID-19 pandemic. Concretely, the full-data random forest model exhibits a predictive out-of-sample accuracy of 53.84%, whereby the respective logistic regression model exhibits a predictive out-of-sample accuracy of 52.74%. In comparison, the random forest and logistic regression models that are trained on data not directly related to COVID-19 exhibit a respective predictive accuracy of 50.87% and 50.56%. However, we

Electronic copy available at: https://ssrn.com/abstract=4516371

still find that all employed models significantly outperform a random classifier. A feature importance analysis of the best model reveals that their predictive power is not concentrated on a single COVID-19-related feature type but spreads over multiple different features. Analyzing the regional differences, we find that features that reflect the pandemic development in the US tend to be more relevant than international developments, which may be explained by the fact that companies in the S&P 500 still generate a majority of their revenue domestically.

References

[1] Badar Nadeem Ashraf. Stock markets' reaction to Covid-19: Moderating role of national culture. *Finance Research Letters*, 41(August 2020): 101857, 2021. ISSN 15446123. doi: 10.1016/j.frl.2020.101857. URL https://doi.org/10.1016/j.frl.2020.101857.

[2] Scott R Baker, Nicholas Bloom, Steven J Davis, and Kyle J Kost. Policy news and stock market volatility. Technical report, National Bureau of Economic Research, 2019.

[3] L Breiman, J H Friedman, R Olshen, and C J Stone. *Classification and Regression Trees*. Routledge, 1984.

[4] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[5] Mark M Carhart. On persistence in mutual fund performance. *The Journal of Finance*, 52(1):57–82, 1997.

[6] Cosmin-octavian Cepoi. Asymmetric dependence between stock market returns and news during COVID-19 financial turmoil. *Finance Research Letters*, 36, 2020. doi: https://doi.org/10.1016/j.frl.2020.101658.

[7] Cetin Ciner. Stock return predictability in the time of COVID-19. *Finance Research Letters*, 38(May 2020):101705, 2021. ISSN 15446123. doi: 10.1016/j.frl.2020.101705. URL https://doi.org/10.1016/j.frl.2020.101705.

[8] Shaen Corbet, Yang Hou, Yang Hu, Brian Lucey, and Les Oxley. Aye Corona! The contagion effects of being named Corona during the COVID-19 pandemic. *Finance Research Letters*, 38(May 2020):

Electronic copy available at: https://ssrn.com/abstract=4516371

101591, 2021. ISSN 15446123. doi: 10.1016/j.frl.2020.101591. URL https://doi.org/10.1016/j.frl.2020.101591.

[9] Aaron Defazio, Francis Bach, and Simon Lacoste-Julien. SAGA: A fast incremental gradient method with support for non-strongly convex composite objectives. *Advances in neural information processing systems*, 27, 2014.

[10] Francis Diebold and Robert Mariano. Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 20(1):134–144, 1994. doi: https://doi.org/10.1198/073500102753410444.

[11] Economic Policy Uncertainty. Daily Infectious Disease Equity Market Volatility Tracker. *https://www.policyuncertainty.com/infectious_EMV.html*, 2022. Last Accessed 2022-05-22.

[12] Eugene F. Fama and Kenneth R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1): 3–56, feb 1993. ISSN 0304405X. doi: 10.1016/0304-405X(93)90023-5.

[13] Eugene F Fama and James D MacBeth. Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, 81(3):607–636, 1973.

[14] Anthony S Fauci, H Clifford Lane, and Robert R Redfield. Covid-19—Navigating the Uncharted. *New England Journal of Medicine*, 382 (13), 2020.

[15] Thomas Fischer and Christopher Krauss. Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2):654–669, 2018. ISSN 03772217. doi: https://doi.org/10.1016/j.ejor.2017.11.054.

[16] Kenneth R. French. Kenneth R. French Data Library. https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html, 2022. Last Accessed 2022-07-21.

[17] Shihao Gu, Bryan Kelly, and Dacheng Xiu. Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies*, 33(5):2223–2273, 2020. doi: https://doi.org/10.1093/rfs/hhaa009.

Electronic copy available at: https://ssrn.com/abstract=4516371

[18] Omair Haroon and Syed Aun R. Rizvi. COVID-19: Media coverage and financial markets behavior—A sectoral inquiry. *Journal of Behavioral and Experimental Finance*, 27:100343, 2020. ISSN 22146369. doi: 10.1016/j.jbef.2020.100343. URL https://doi.org/10.1016/j.jbef.2020.100343.

[19] David W Hosmer Jr, Stanley Lemeshow, and Rodney X Sturdivant. *Applied logistic regression*, volume 398. John Wiley & Sons, 2013.

[20] James K. Jackson. Global Economic Effects of Covid 19. Technical report, Congressional Research Service, 2021.

[21] Johns Hopkins University. COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. https://github.com/CSSEGISandData/COVID-19, 2022.

[22] Christopher Krauss, Xuan Anh Do, and Nicolas Huck. Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research*, 259(2): 689–702, 2017.

[23] Weiqing Li, Fengsheng Chien, Hafiz Waqas Kamran, Talla M. Aldeehani, Muhammad Sadiq, Van Chien Nguyen, and Farhad Taghizadeh-Hesary. The nexus between COVID-19 fear and stock market volatility. *Economic Research-Ekonomska Istrazivanja*, 0(0):1–22, 2021. ISSN 1331677X. doi: 10.1080/1331677X.2021.1914125. URL https://doi.org/10.1080/1331677X.2021.1914125.

[24] NT Pramathesh Mishra, Sabya Sachi Das, Shalini Yadav, Wasim Khan, Mohd Afzal, Abdullah Alarifi, El Refaie Kenawy, Mohammed Tahir Ansari, Md Saquib Hasnain, and Amit Kumar Nayak. Global impacts of pre- and post-COVID-19 pandemic: Focus on socio-economic consequences. *Sensors International*, 1(July), 2020. ISSN 26663511. doi: 10.1016/j.sintl.2020.100042.

[25] Paresh Kumar Narayan, Dinh Hoang Bach Phan, and Guangqiang Liu. COVID-19 lockdowns, stimulus packages, travel bans, and stock returns. *Finance Research Letters*, 38(June 2020):101732, 2021. ISSN 15446123. doi: 10.1016/j.frl.2020.101732. URL https://doi.org/10.1016/j.frl.2020.101732.

Electronic copy available at: https://ssrn.com/abstract=4516371

[26] Maria Nicola, Zaid Alsafi, Catrin Sohrabi, Ahmed Kerwan, Ahmed Al-Jabir, Christos Iosifidis, Maliha Agha, and Riaz Agha. The socio-economic implications of the coronavirus pandemic (COVID-19): A review. *International Journal of Surgery*, 78(March):185–193, 2020. ISSN 17439159. doi: 10.1016/j.ijsu.2020.04.018. URL https://doi.org/10.1016/j.ijsu.2020.04.018.

[27] Enrico Onali. COVID-19 and Stock Market Volatility. *SSRN Electronic Journal*, pages 1–24, 2020. doi: 10.2139/ssrn.3571453.

[28] Peterson K. Ozili and Thankom G. Arun. Spillover of COVID-19: impact on the Global Economy. *SSRN*, 2020. doi: https://dx.doi.org/10.2139/ssrn.3562570. URL https://ssrn.com/abstract=3562570.

[29] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, and Others. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[30] Anh Viet Pham, Christofer Adrian, Mukesh Garg, Soon Yeow Phang, and Cameron Truong. State-level COVID-19 outbreak and stock returns. *Finance Research Letters*, 43(September 2020):102002, 2021. ISSN 15446123. doi: 10.1016/j.frl.2021.102002. URL https://doi.org/10.1016/j.frl.2021.102002.

[31] Hannah Ritchie, Edouard Mathieu, Lucas Rodés-Guirao, Cameron Appel, Charlie Giattino, Esteban Ortiz-Ospina, Joe Hasell, Bobbie Macdonald, Diana Beltekian, and Max Roser. Coronavirus pandemic (covid-19). https://ourworldindata.org/coronavirus, 2020. Last Accessed 2022-05-21.

[32] Afees A. Salisu and Xuan Vinh Vo. Predicting stock returns in the presence of COVID-19 pandemic: The role of health news. *International Review of Financial Analysis*, 71(June):101546, 2020. ISSN 10575219. doi: 10.1016/j.irfa.2020.101546. URL https://doi.org/10.1016/j.irfa.2020.101546.

[33] Afees A. Salisu, Lateef Akanni, and Ibrahim Raheem. The COVID-19 global fear index and the predictability of commodity price returns. *Journal of Behavioral and Experimental Finance*, 27:100383, 2020. ISSN

Electronic copy available at: https://ssrn.com/abstract=4516371

22146369. doi: 10.1016/j.jbef.2020.100383. URL https://doi.org/10.1016/j.jbef.2020.100383.

[34] Howard Silverblatt. SP 500® 2018: Global Sales. *Working Paper*, 2019. URL https://www.spglobal.com/spdji/en/documents/research/sp-500-global-sales-2018.pdf. Last Accessed 2022-07-21.

[35] Sowmya Subramaniam and Madhumita Chakraborty. COVID-19 fear index: does it matter for stock market returns? *Review of Behavioral Finance*, 13(1):40–50, 2021. ISSN 19405987. doi: 10.1108/RBF-08-2020-0215.

[36] University of Pennsylvania. Wharton Research Data Services. https://wrds-www.wharton.upenn.edu/, 2022. Last Accessed 2022-07-21.

[37] Quan Jing Wang, Di Chen, and Chun Ping Chang. The impact of COVID-19 on stock prices of solar enterprises: A comprehensive evidence based on the government response and confirmed cases. *International Journal of Green Energy*, 18(5):443–456, 2021. ISSN 15435083. doi: 10.1080/15435075.2020.1865367. URL https://doi.org/10.1080/15435075.2020.1865367.

[38] Worldbank. COVID-19 Finance Sector Related Policy Responses. https://datacatalog.worldbank.org/search/dataset/0037999/COVID-19-Finance-Sector-Related-Policy-Responses, 2022. Last Accessed 2022-05-21.

Electronic copy available at: https://ssrn.com/abstract=4516371